# Medical Decision-Making in the Physician Hierarchy: A Pilot Pedagogical Evaluation

John Rosasco[1], Michele L McCarroll[1] iD, M David Gothard[2], Jerry Myers[3], Patrick Hughes[4], Alan Schwartz[5], Richard L George[6,7] and Rami A Ahmed[8]

[1]College of Osteopathic Medicine, Pacific Northwest University of Health Sciences, Yakima, WA, USA. [2]Biostats, Inc., East Canton, OH, USA. [3]HRP Cross Cutting Computational Modeling Project, NASA John H. Glenn Research Center, Cleveland, OH, USA. [4]Department of Emergency Medicine, Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, USA. [5]Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA. [6]Department of Surgery, Trauma Program, Summa Health System —Akron Campus, Akron, OH, USA. [7]Department of Surgery, Northeast Ohio Medical University, Rootstown, OH, USA. [8]Department of Emergency Medicine, School of Medicine, Indiana University, Indianapolis, IN, USA.

## ABSTRACT

**PURPOSE:** Recently, the American College of Graduate Medical Education included medical decision-making as a core competency in several specialties. To date, the ability to demonstrate and measure a pedagogical evolution of medical judgment in a medical education program has been limited. In this study, we aim to examine differences in medical decision-making of physician groups in distinctly different stages of their postgraduate career.

**METHODS:** The study recruited physicians with a wide spectrum of disciplines and levels of experience to take part in 4 medical simulations divided into 2 categories, abdominal pain (biliary colic [BC] and renal colic [RC]) or chest pain (cardiac ischemia with ST-segment elevation myocardial infarction [STEMI] and pneumothorax [PTX]). Evaluation of medical decision-making used the Medical Judgment Metric (MJM). The targeted selection criteria for the physician groups are administrative physicians (APs), representing those with the most experience but whose current duties are largely administrative; resident physicians (RPs), those enrolled in postgraduate medical or surgical training; and mastery level physicians (MPs), those deemed to have mastery level experience. The study measured participant demographics, physiological responses, medical judgment scores, and simulation time to case resolution. Outcome differences were analyzed using Fisher exact tests with post hoc Bonferroni-adjusted $z$ tests and single-factor analysis of variance F tests with post hoc Tukey honestly significant difference, as appropriate. The significance threshold was set at $P < .05$. Effect sizes were determined and reported to inform future studies.

**RESULTS:** A total of n = 30 physicians were recruited for the study with n = 10 participants in each physician group. No significant differences were found in baseline demographics between groups. Analysis of simulations showed a significant ($P = .002$) interaction for total simulation time between groups RP: 6.2 minutes ($\pm 1.58$); MP: 8.7 minutes ($\pm 2.46$); and AP: 10.3 minutes ($\pm 2.78$). The AP MJM scores, 12.3 ($\pm 2.66$), for the RC simulation were significantly ($P = .010$) lower than the RP 14.7 ($\pm 1.15$) and MP 14.7 ($\pm 1.15$) MJM scores. Analysis of simulated patient outcomes showed that the AP group was significantly less likely to stabilize the participant in the RC simulation than MP and RP groups ($P = .040$). While not significant, all MJM scores for the AP group were lower in the BC, STEMI, and PTX simulations compared with the RP and MP groups.

**CONCLUSIONS:** Physicians in distinctly different stages of their respective postgraduate career differed in several domains when assessed through a consistent high-fidelity medical simulation program. Further studies are warranted to accurately assess pedagogical differences over the medical judgment lifespan of a physician.

**KEYWORDS:** Skill retention, knowledge retention, clinical judgment, medical judgment, decision-making, simulation, medical education

## Introduction

Pedagogy is defined as the art, science, or profession of teaching medical learning based on learning theory in which there is a transmittal of knowledge.[1,2] In medical education, levels of pedagogical transitions include undergraduate medical education, graduate medical education, and continuing medical education (CME), where evidenced-based and scientifically based knowledge is applied to ensure quality performance.[3] Over time, efforts to assess a physician's skill and knowledge retention, as well as examine physician competency, used standard training checkpoints such as state licensing requirements, national board exams, CME requirements for specialties, and

board recertification standards. However, accurately measuring and determining competency in clinical practice from these assessments remains a challenge.[4]

Despite set standards for lifelong medical education, the ability to ensure retention of expert medical decision-making, skill, and knowledge remains highly variable and challenged by the limited ability to quantitatively assess such skills.[5] Medical decision-making describes the ability to build ties between data obtained from history, physical examination, imaging results, and laboratory studies to formulate assessments and plans for patients.[6] Good medical decision-making leads to improved outcomes for patients, whereas poor medical decision-making is associated with morbidity and mortality in patients.[7] To underscore the importance of testing medical decision-making, the American College of Graduate Medical Education added medical decision-making to their list of core competencies that every physician should possess.[8-10] As such, promoting strong medical judgment skills is a unique challenge in the evaluation of physician training, particularly as medical education continues to develop. Thus, a need exists for standardized criteria to evaluate medical decision-making, not only among medical students but also among postgraduate physicians.[11,12]

To meet the need for clinicians to possess strong medical decision-making, medical education measured competency and judgment by integrating key competencies in medical simulation, problem-based learning, and case-based discussions.[13-16] However, a reproducible simulation environment married to a consistent quantitative assessment metric that allows measurement of medical decision-making remains elusive.

Simulation laboratories may provide the safe, reproducible environment required to accurately measure decision-making. Simulation was designed to provide a reproducible training environment and has its roots in aerospace sciences and engineering. Flight simulation, since the development and implementation of the early flight sims in the 1930s, endures as an integral part of both pilot and astronaut training due to its ability to test technical competency as well as provide participants with applicable situational experience.[17] Simulation training provides pilots and astronauts with the skills to adapt to unfamiliar scenarios during performance of their duties. Medicine bears similarities in the potential for unpredictable situations to arise quickly and requires that physicians be able to make competent judgments to achieve positive outcomes. To improve the probability of effective performance, both industries have widely implemented simulation to provide experience and training in complex or rare situations.[18]

High-fidelity simulations followed by effective debriefing in the simulation laboratory have caused simulation to emerge as a high-impact teaching tool with evidence to suggest performance may transfer to improved performance clinically.[19-21] However, there is currently limited evidence that exists to suggest that outcomes in simulated scenarios correlate to patient outcomes during live patient care.[20] Macgahie Medical education's widespread adoption of simulation does, however,

provide an ideal and reproducible environment to evaluate medical decision-making. Despite this ideal environment, implementation of a unified metric for measuring competency that minimizes assessment variability remains elusive. The need for a consistent assessment tool led to the development of the Medical Judgment Metric (MJM).[22] The metric creates a standardized measurement of the complex multidimensional thinking required of strong clinicians. By design, the MJM measures decision-making in challenging environments and differentiates between physicians who can perform well on the earlier stated training checkpoints from those who excel at the most difficult part of their responsibilities, making decisions for patients in critical situations. Previous work in the development of the MJM involved measuring the decision-making of National Aeronautics and Space Administration (NASA) spaceflight crew medical officers against their nonmedically trained peers.[23] In this investigation, we seek to isolate and evaluate the medical decision-making of the 3 different physician groups at different stages of medical practice.

## Methods

### Participants

The Summa Health Institutional Review Board and NASA Johnson Space Center Committee for the Protection of Human Subjects were responsible for approving the study protocol. Detailed procedures of recruitment and rationale for participant selection criteria were reported in previous publications by Ahmed et al[22] and McCarroll et al.[23] Overall, participants were recruited by fliers, emails, and department meetings at local hospitals and universities in Northeast Ohio during a 6-month timeframe. Participant recruitment focused on strict screening criteria to ensure adequate sampling across 3 physician groups to represent various levels of pedagogical transitions: administrative physicians (APs; n = 10), mastery level physicians (MPs; n = 10), and resident physicians (RPs; n = 10). Specifically, the researchers recruited participants based on the current NASA astronaut pool demographic data to create an appropriate analogue population for the study. APs are considered the most experienced group but those whose current duties are largely administrative, performing a maximum of 8 hours of clinical care per week; MPs are attending physicians who have completed postgraduate education and whose responsibilities are entirely clinical. Resident physicians were selected from those who are in the third or fourth year of their postgraduate year education and, thus, are in the final stages of medical training.

### Simulations

Detailed procedures of each simulation were reported in previous publications by Ahmed et al[22] and McCarroll et al.[23] In summary, each group of participants (APs, MPs, and RPs) were scheduled 1 practice simulated scenario and 4 tested scenarios: Abdominal pain scenarios were biliary colic (BC) and renal

colic (RC) simulations, whereas chest pain scenarios were car- diac ischemia with ST-segment elevation myocardial infarc- tion (STEMI) and tension pneumothorax (PTX). Each participant was assigned to complete the scenarios in a random order decided by random numbers generator software in Microsoft Excel V.2007 (Redmond, Washington, USA). The simulations were then performed in a mature medical simula- tion laboratory in an American College of Surgeons verified Level I trauma center in the United States. The simulation laboratories used the METI Emergency Care Simulator man- nequin simulation technology with HPS6 software (Medical Education Technologies, Saint-Laurent, Quebec, Canada).

Participants were informed they had 15 minutes to treat the simulated patient, and all simulations were timed as well as audio and video recorded in their entirety. Participants' heart rates were captured using a heart rate monitor worn by the participant (Polar Heart Rate Monitor; Polar Electro, Lake Success, New York, USA) where baseline and maximum heart rates were recorded during each simulation. During a 5-minute break following each simulation, NASA Task Load Index (NASA TLX) surveys were performed by the participant and a resting heart rate was obtained for each participant.

Medical decision-making in each of the 4 scenarios was evaluated using a scenario-specific critical action checklist sim- ilar to that used in prior MJM experiments, a categorical deter- mination of patient outcome (loss of function, loss of life, or stabilized), as well as the MJM. The MJM uses 4 clinical domains: history and physical, diagnostics, interpretation, and management to assess medical decision-making. Four review- ers were used to score participants using the MJM, each with postgraduate training in either emergency medicine, general surgery with surgical critical care fellowship training, and/or fellowship training in medical simulation in addition to being trained to use the MJM. The raters were blinded to both the participant's name and cohort. Inter-rater operability of the MJM by reviewers was previously performed.[22]

### Statistical analysis

Statistical comparison of demographics and the main variables of physiological responses, NASA TLX, and MJM medical decision-making scores per simulated scenario were performed using nonparametric *t* tests with Fisher exact test for categori- cal variables or the Mann-Whitney *U* test for numerical vari- ables. *P* values for physiologic scores were from repeated measures analysis of variance (ANOVA), and *P* values for total sim time were from single-factor ANOVA with post hoc Tukey HSD tests. Analysis of MJM scores was performed using sin- gle-factor ANOVA F tests with post hoc Tukey HSD tests. Outcome differences were analyzed using Fisher exact tests with post hoc Bonferroni-adjusted *z* tests, as appropriate. The significance threshold was set at $P < .05$. Effect sizes were determined and reported to inform future studies.[24]

## Results

A total of n = 30 physicians were recruited for the study. Participant demographics and characteristics are summa- rized in Table 1. All participants had obtained a doctoral degree in medicine or osteopathic medicine (MD or DO), but specialty training varied within each group. Age charac- teristics are younger for the RP group and older for the AP group. Race distribution within each group was mostly white. In keeping with prior experiments with the MJM, sex between all the groups was maintained at 60% men and 40% women. Most participants identified with training in emer- gency medicine (n = 14); general surgery (n = 5) as well as obstetrics and gynecology (n = 4) were the next highest disci- plines represented.

Participant physiologic data are summarized in Table 2. No significant difference was observed in the NASA TLX score between groups (P = .161), though anecdotally the MP has the lowest mean NASA TLX scores in each of the simulations. Baseline heart rate mean was not significantly different among groups (P = .907), and maximum heart rate mean showed a similar anecdotal trend of MP group having lower maximum heart rates but was still not considered significant (P = .283). While the mean total simulation time was not significantly dif- ferent among groups (P = .204), a significant time × group interaction was present (P = .033), hence the between-group comparisons at each time period. A significant difference between groups was noted in total simulation time (P = .002) for the RC medical simulation with the RP group 6.2 minutes (±1.58), MP group 8.7 minutes (±2.46), and AP group 10.3 minutes (±2.78).

Table 3 summarizes MJM scores. There was a statistical (P = .010) difference between the AP group relative to the other 2 groups for the RC simulations of MJM judgment scores: RP = 14.7 score (±1.15), MP = 14.7 score (±1.48) and AP = 12.3 score (±2.66). Finally, a statistical (P = .040) differ- ence was noted in the RC simulation in outcomes between groups with the AP group performing poorly compared with the other 2 groups, (RP: 0 loss of function, 0 loss of life, 10 stabilized; MP: 1 loss of function, 0 loss of life, 9 stabilized; AP: 3 loss of function, 2 loss of life, 5 stabilized; Table 4). In all other medical simulation scenarios and outcomes, no other sig- nificant differences were observed.

The effect sizes of these data are NASA TLX: 0.127 ($\eta^2$), resting HR: 0.007 ($\eta^2$), max HR: 0.089 ($\eta^2$), and total sim time: 0.111 ($\eta^2$). These data reflect a small effect on resting HR and medium effect approaching large effect differences on max HR, sim time, and NASA TLX between the 3 groups. Other metrics were reported in previously published articles.[22,23]

## Discussion

The present study demonstrated significant differences in medical decision-making between resident, board-certified, and AP groups in a controlled medical simulation of RC. The

**Table 1.** Demographics.

| CHARACTERISTIC | PGY LEVEL 3 OR 4 (n = 10) | BOARD-CERTIFIED PHYSICIAN (n = 10) | ADMINISTRATIVE PHYSICIAN (n = 10) |
|---|---|---|---|
| Age (years) | | | |
| Mean (SD) | 28.6 (1.43) | 42.3 (10.83) | 59.6 (7.78) |
| Median (IQR) | 28.5 (27.8-29) | 38.5 (33.8-55) | 61.5 (51.8-66) |
| Min-Max | 27-32 | 28-57 | 49-70 |
| Sex, n (%) | | | |
| Women | 4 (40%) | 4 (40%) | 4 (40%) |
| Men | 6 (60%) | 6 (60%) | 6 (60%) |
| Race, n (%) | | | |
| Asian | 1 (10%) | 0 | 1 (10%) |
| Asian/white | 2 (20%) | 0 | 0 |
| Black | 0 | 1 (10%) | 0 |
| White | 7 (70%) | 9 (90%) | 9 (90%) |
| Specialty, n (%) | | | |
| Emergency medicine | 4 (40%) | 7 (70%) | 3 (30%) |
| General surgery | 4 (40%) | 1 (10%) | 0 |
| OB/GYN | 2 (20%) | 1 (10%) | 1 (10%) |
| Radiology | 0 | 1 (10%) | 0 |
| Family medicine | 0 | 0 | 1 (10%) |
| Internal medicine and pediatrics | 0 | 0 | 1 (10%) |
| Internal medicine | 0 | 0 | 1 (10%) |
| Internal medicine and infectious diseases | 0 | 0 | 1 (10%) |
| OB/GYN and preventative | 0 | 0 | 1 (10%) |
| Pathology | 0 | 0 | 1 (10%) |

Abbreviations: IQR, interquartile range; OB/GYN, obstetrics and gynecology; PGY, postgraduate year.

study implemented several metrics used to measure medical decision-making such as the MJM, NASA TLX, and physiological responses to simulations. Analysis of the results of the MJM showed a significant drop in the MJM scores for the AP group as well as significantly poorer prognostic indications compared with early and mid-career physicians. Of note, the AP group had nonsignificant decreases in most outcomes measuring medical decision-making and responses to medical trauma across all simulated scenarios. These results warrant further exploration regarding the numerous and varied duties asked of APs, especially as it relates to emergency scenarios.

Several differences exist in the daily responsibilities of the 3 physician groups studied in this investigation. In general, AP physicians perform less than 8 hours of clinical work per week, compared with RP physicians who work up to 80 hours a week

performing clinical duties, and MP physicians who spend on average 30 to 45 hours seeing patients each week.[25] Not only do the work hours and years of experience make up a physician's capability for medical decision-making, medical specialties differ widely of responsibilities, training requirements, and hours worked. As such, certain specialties may have decreased exposure to acute care cases. While the data of work hours are clear, the pedagogical evolution of medical education, credentialing requirements, and judgment across the career of a physician is difficult to measure. Cook et al[26] concluded that Kane framework is a valid means of measuring a medical professional's decision-making because the pillars include interpreting clinical laboratory tests, procedural skills competency, and qualitative (narrative) evaluation. The current study applied 4 different medical simulations in which each simulation tested the

**Table 2.** Physiological Data.

| SIMULATION NUMBER/METRIC/ STATISTIC | PGY LEVEL 3 OR 4 (n = 10) | BOARD-CERTIFIED PHYSICIAN (n = 10) | ADMINISTRATIVE PHYSICIAN (n = 10) | BETWEEN GROUPS (*P* VALUE) |
|---|---|---|---|---|
| Between-group effect | | | | |
| NASA TLX score | | | | .161 |
| Resting heart rate | | | | .907 |
| Maximum heart rate | | | | .283 |
| Total sim time | | | | .204 |
| Renal colic | | | | |
| NASA TLX score mean (SD) | 43.7 (16.93) | 37.6 (13.91) | 54.4 (21.59) | |
| Resting heart rate mean (SD) | 77.5 (7.86) | 76.0 (14.78) | 75.3 (13.73) | |
| Maximum heart rate mean (SD) | 99.4 (12.79) | 88.3 (14.48) | 93.2 (15.50) | |
| Total sim time mean (SD) | 6.2 (1.58) | 8.7 (2.46) | 10.3 (2.78) | **.002** |
| Biliary colic | | | | |
| NASA TLX score mean (SD) | 53.6 (25.76) | 38.3 (20.23) | 52.2 (31.11) | |
| Resting heart rate mean (SD) | 77.9 (8.05) | 76.5 (9.85) | 77.1 (12.09) | |
| Maximum heart rate mean (SD) | 95.7 (9.26) | 87.2 (14.49) | 90.1 (15.39) | |
| Total sim time mean (SD) | 8.6 (2.78) | 7.6 (1.94) | 9.2 (2.65) | .339 |
| STEMI | | | | |
| NASA TLX score mean (SD) | 47.2 (25.71) | 43.2 (25.57) | 54.3 (26.04) | |
| Resting heart rate mean (SD) | 77.9 (9.97) | 74.1 (11.13) | 76.2 (12.40) | |
| Maximum heart rate mean (SD) | 94.6 (12.42) | 87.2 (11.78) | 90.4 (15.41) | |
| Total sim time mean (SD) | 7.9 (2.93) | 7.9 (1.40) | 8.0 (2.97) | .994 |
| Pneumothorax | | | | |
| NASA TLX score mean (SD) | 42.1 (19.43) | 28.6 (11.44) | 50.5 (30.04) | |
| Resting heart rate mean (SD) | 75.3 (9.29) | 73.6 (13.10) | 75.3 (13.01) | |
| Maximum heart rate mean (SD) | 93.8 (8.09) | 84.3 (13.98) | 87.9 (13.54) | |
| Total sim time mean (SD) | 7.3 (2.13) | 7.2 (1.74) | 8.0 (2.36) | .626 |

Bold vaule signifies 0.002. *P* values from between-group factor of repeated measures ANOVA. For total sim time, a significant (*P* = .033) time × group interaction term was present, hence the between-group comparisons at each time period. *P* value for total sim time from single-factor ANOVA with the post hoc Tukey HSD test. PGY 3 or 4 differed significantly at sim 1 from the administrative group.
Abbreviations: ANOVA, analysis of variance; NASA TLX, National Aeronautics and Space Administration Task Load Index; PGY, postgraduate year; STEMI, ST-segment elevation myocardial infarction.

medical professional in Kane pillars while simultaneously applying a new medical judgment tool (MJM).[27]

The AP group, versus the RP and MP groups, had a significantly wider variety of specialty training and included 6 specialties not seen in the other physician groups. It is also plausible that the AP group's career focus may have shifted, potentially resulting in decreased clinical acumen. Kneebone and Nestel[28] describe the varying familiarity with simulation, which may correlate with the age of physicians (as in the AP group) as simulation in medical education did not become widespread until the early 21st century. Dellinger et al[29] and Goldberg

et al[30] have reported that as physician's age, cognitive ability declines 20% between age 40 and 75 years. Furthermore, as physicians approach retirement, they have more emotional exhaustion after shifts, less ability to handle stress and heavy workloads, and some degree of physical and cognitive decline.[29,30] To combat this decline, Skowronski and Peisah suggested several strategies, including a reduced exposure to acute crisis intervention for intensivists, with an increased focus on mentoring, teaching, administration, and research.[28,31] As physicians enter the later years in their profession, they tend to receive more support and may have more resources they rely on

**Table 3.** Analysis of Medical Judgment Metric Scores.

| SIMULATION/ STATISTIC | PGY LEVEL 3 OR 4 (n = 10) | BOARD-CERTIFIED PHYSICIAN (n = 10) | ADMINISTRATIVE PHYSICIAN (n = 10) | *P* VALUE |
|---|---|---|---|---|
| Renal colic | | | | **.010** |
| Mean (SD) | 14.7 (1.15) | 14.7 (1.48) | 12.3 (2.66) | |
| Median (IQR) | 14.8 (14.2-15.7) | 15.4 (13.5-15.7) | 12.4 (10.6-14.5) | |
| Min-Max | 12.2-15.8 | 12-16 | 8.2-16 | |
| Biliary colic | | | | .269 |
| Mean (SD) | 14.5 (1.16) | 14.2 (2.68) | 13.0 (2.25) | |
| Median (IQR) | 14.8 (13.5-15.2) | 15.5 (12.6-15.9) | 13.1 (11.5-14.9) | |
| Min-Max | 12.5-16 | 7.7-16 | 8.8-16 | |
| STEMI | | | | .175 |
| Mean (SD) | 14.2 (1.93) | 14.1 (2.71) | 12.1 (3.33) | |
| Median (IQR) | 14.6 (12.7-16) | 15.7 (10.5-16) | 11.8 (9.5-15.4) | |
| Min-Max | 11-16 | 9.5-16 | 7-16 | |
| Pneumothorax | | | | .336 |
| Mean (SD) | 14.4 (1.75) | 14.8 (1.12) | 13.5 (2.53) | |
| Median (IQR) | 15.3 (13.2-15.9) | 14.8 (14-15.8) | 14.3 (11.5-15.7) | |
| Min-Max | 11-16 | 12.7-16 | 9-15.8 | |

Bold vaule signifies 0.010. *P* values from single-factor ANOVA F test with the post hoc Tukey HSD test. For the renal colic simulation, the administrative study group was statistically distinct ($P < .001$) in their MJPM scores relative to the other 2 study groups which were statistically indistinguishable.
Abbreviations: ANOVA, analysis of variance; IQR, interquartile range; PGY, postgraduate year; STEMI, ST-segment elevation myocardial infarction.

**Table 4.** Categorical Analysis of Outcomes.

| SIMULATION/ OUTCOME | PGY LEVEL 3 OR 4 (n = 10) | BOARD-CERTIFIED PHYSICIAN (n = 10) | ADMINISTRATIVE PHYSICIAN (n = 10) | *P* VALUE |
|---|---|---|---|---|
| Renal colic | | | | **.040** |
| Loss of function | 0 | 1 (10%) | 3 (30%) | |
| Loss of life | 0 | 0 | 2 (20%) | |
| Stabilized | 10 (100%) | 9 (90%) | **5 (50%)** | |
| Biliary colic | | | | .457 |
| Loss of function | 2 (20%) | 1 (10%) | 3 (30%) | |
| Loss of life | 0 | 1 (10%) | 2 (20%) | |
| Stabilized | 8 (80%) | 8 (80%) | 5 (50%) | |
| STEMI | | | | .610 |
| Loss of function | 1 (10%) | 0 | 2 (20%) | |
| Loss of life | 4 (40%) | 3 (30%) | 4 (40%) | |
| Stabilized | 5 (50%) | 7 (70%) | 4 (40%) | |
| Pneumothorax | | | | .877 |
| Loss of function | 0 | 1 (10%) | 2 (20%) | |
| Loss of life | 2 (20%) | 1 (10%) | 1 (10%) | |
| Stabilized | 8 (80%) | 8 (80%) | 7 (70%) | |

*P* values from Fisher exact tests with post hoc Bonferroni-adjusted *z* tests. Bold values indicate values with significant Bonferroni-adjusted *z* tests.
Abbreviations: PGY, postgraduate year; STEMI, ST-segment elevation myocardial infarction.

that assist in their decision-making capacity leading to a shift away from clinical responsibilities.[32]

Limitations to the study are vast. While the physician group variations may offer explanation for the observed differences, they warrant further exploration as is the goal of the primary study was not to examine the reasons for differences in the decision-making capacity among APs or their peers. This study was set to examine potential quantitative differences based on validated and nonvalidated tools to determine differences in medical decision-making among physician groups with varying levels of training. The use of the MJM (nonvalidated) in a limited convenience sample of physicians was intended to continue to establish its efficacy as quantitative measurement tool for future studies. Based on the effect sizes from this study, future studies can ensure proper sample sizes to offer conclusions about medical decision-making using the MJM. In the future, the MJM, along with more patient shared decision-making simulation studies, may be a valuable tool to differentiate physician medical decision-making stages over a career and generate hypotheses about differences.

Previous work involving the MJM established the use of the MJM as a potential tool to determine the competency of spaceflight crew medical officers and their decision-making compared with nonmedically trained spaceflight crew members. In this study, we continue to define a role for the MJM as a tool that can be implemented to assess competency in medical decision-making among physicians of different trainings levels. When combined with professionally written and implemented simulation scenarios, critical action checklists, as well as dedicated simulation resources and team members, the MJM can provide a useful tool to quantify physician's decision-making. Using the MJM to assess competencies among a larger cohort of peers, particularly within the same level of training and specialty, is the next step toward establishing the MJM as a reliable quantitative tool that is able to effectively measure differences in decision-making among physicians and thus allow improved feedback following simulations eventually providing improvements in clinical care and patient outcomes.

## Authors' Note

## Acknowledgements

## Author Contributions

All authors met the International Committee of Medical Journal Editors definition of authorship and intellectual contribution.

## ORCID iD

Michele L McCarroll ⓘD https://orcid.org/0000-0003-0950-8882

## REFERENCES

1.  Merriam-Webster Dictionary. https://www.merriam-webster.com/dictionary/pedagogy. Accessed December 9, 2018.
2.  The Free Dictionary. https://medical-dictionary.thefreedictionary.com/pedagogy. Accessed December 9, 2018.
3.  American College of Graduate Medical Education (ACGME). http://www.accme.org/accreditation-rules/policies/cme-content-definition-and-examples. Accessed December 9, 2018.
4.  Schwartz A. Medical decision making and medical education: challenges and opportunities. *Perspect Biol Med*. 2011;54:68-74.
5.  Murray DJ, Freeman BD, Boulet JR, Woodhouse J, Fehr JJ, Klingensmith ME. Decision making in trauma settings: simulation to improve diagnostic skills. *Simul Healthc*. 2015;10:139-145.
6.  Sox Harold C. *Medical Decision Making*. Chichester, UK: John Wiley & Sons; 2013.
7.  Roter D. The patient-physician relationship and its implications for malpractice litigation. *J Health Care L & Pol'y*. 2006;9:304-314.
8.  Lyss-Lerman P, Teherani A, Aagaard E, Loeser H, Cooke M, Harper GM. What training is needed in the fourth year of medical school? views of residency program directors. *Acad Med*. 2009;84:823-829.
9.  Frischknecht AC, Boehler ML, Schwind CJ, et al. How prepared are your interns to take calls? Results of a multi-institutional study of simulated pages to prepare medical students for surgery internship. *Am J Surg*. 2014;208:307-315.
10. Chamberland M, Mamede S, St-Onge C, Setrakian J, Schmidt HG. Does medical students' diagnostic performance improve by observing examples of self-explanation provided by peers or experts? *Adv Health Sci Educ Theory Pract*. 2015;20:981-993.
11. Cohen JJ. Medical education in an era of health-care reform. *Perspect Biol Med*. 2011;54:61-67.
12. Croskerry P, Petrie D, Reilly J, Tait G. Deciding about fast and slow decisions. *Acad Med*. 2014;89:197-200.
13. Okuda Y, Bryson EO, DeMaria S Jr, et al. The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med*. 2009;76:330-343.
14. Onyon C. Problem-based learning: a review of the educational and psychological theory. *Clin Teach*. 2012;9:22-26.
15. McLean SF. Case-based learning and its application in medical and health-care fields: a review of worldwide literature [published online ahead of print April 27, 2016]. *J Med Educ Curric Dev*. doi:10.4137/JMECD.S20377.
16. Morrow CE, Reed VA, Eliassen MS, Imset I. Shared decision making: skill acquisition for year III medical students. *Fam Med*. 2011;43:721-725.
17. Page RL. *Brief History of Flight Simulation*. Sydney, NSW, Australia: The Sim-TecT 2000 Organising and Technical Committee; 2000.
18. Lateef F. Simulation-based learning: just like the real thing. *J Emerg Trauma Shock*. 2010;3:348-352. doi:10.4103/0974-2700.70743.
19. Wayne DB, Didwania A, Feinglass J, Fudala M, Barsuk J, McGaghie W. Simulation based education improves quality of care during cardiac arrest team responses at an academic teaching hospital: a case-control study. *Chest*. 2008;133:56-61.
20. McGaghie WC, Draycott TJ, Dunn WF, Lopez CM, Stefanidis D. Evaluating the impact of simulation on translational patient outcomes. *Simul Healthc*. 2011;6:S42-S47.
21. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med*. 2009;37:2697-2701.
22. Ahmed RA, McCarroll ML, Schwartz A, et al. Development, validation, and implementation of a medical judgment metric [published online ahead of print June 19, 2017]. *MDM Pol Prac*. doi:10.1177/2381468317715262.
23. McCarroll ML, Ahmed RA, Schwartz A, et al. Medical judgement analogue studies with applications to spaceflight crew medical officers. *BMJ Simul Tech Enhan Learn*. 2017;3:163-168.
24. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
25. American College of Graduate Medical Education (ACGME). https://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements. Accessed May 1, 2019.
26. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49:560-575. doi:10.1111/medu.12678.
27. Petterson SM, Phillips RL Jr, Bazemore AW, Koinis GT. *Unequal Distribution of the U.S. Primary Care Workforce*. Washington, DC: Robert Graham Center; 2013.
28. Kneebone R, Nestel D. Learning clinical skills: the place of simulation and feedback. *Clin Teac*. 2005;2:86-90.

29. Dellinger EP, Pellegrini CA, Gallagher TH. The aging physician and the medical profession: a review. *JAMA Surg*. 2017;152:967-971. doi:10.1001/jamasurg.2017.2342.

30. Goldberg R1, Thomas H, Penner L. Issues of concern to emergency physicians in pre-retirement years: a survey. *J Emerg Med*. 2011;40:706-713. doi:10.1016/j.jemermed.2009.08.020.

31. Skowronski GA, Peisah C. The greying intensivist: ageing and medical practice—everyone's problem. *Med J Aust*. 2012;196:505-507.

32. Smith F, Lachish S, Goldacre MJ, Lambert TW. Factors influencing the decisions of senior UK doctors to retire or remain in medicine: national surveys of the UK-trained medical graduates of 1974 and 1977. *BMJ Open*. 2017;7:e017650. doi:10.1136/bmjopen-2017-017650.